

WS

中华人民共和国卫生行业标准

WS/T 408—2024

代替WS/T 408—2012, WS/T 416—2013, WS/T 420—2013 和 WS/T 492—2016

定量检验程序分析性能验证指南

Guideline for the verification of analytical performance of quantitative examination
procedures

2024 - 05 - 09 发布

2024 - 11 - 01 实施

中华人民共和国国家卫生健康委员会 发布

前 言

本标准为你推荐性标准。

本标准代替WS/T 408—2012《临床化学设备线性评价指南》、WS/T 416—2013《干扰试验指南》、WS/T 420—2013《临床实验室对商品定量试剂盒分析性能的验证》和WS/T 492—2016《临床检验定量测定项目精密度与正确度性能验证》等四项标准。本标准主要技术变化如下：

- 更改了标准名称，本标准替代的四项标准统称为“定量检验程序分析性能验证指南”；
- 更改了范围，标准用途限于临床实验室分析性能验证；
- 重新组织了标准内容，集中描述原则性要求和共性技术问题，对各项验证进行结构化描述；
- 整合了重复内容，更改了部分方案设计和数据处理，简化了部分操作和数据处理细节；
- 更改了验证实验结果解释，以临床需要的性能水平做主要判读依据，利用统计学检验协助分析判断。

本标准由国家卫生健康标准委员会临床检验标准专业委员会负责技术审查和技术咨询，由国家卫生健康委医疗管理服务指导中心负责协调性和格式审查，由国家卫生健康委员会医政司负责业务管理、法规司负责统筹管理。

本标准起草单位：北京医院/国家卫生健康委临床检验中心、民航总医院、中国医学科学院北京协和医院、首都医科大学附属北京同仁医院、深圳市罗湖医院集团、广东省中医院、同济大学附属杨浦医院。

本标准主要起草人：陈文祥、张传宝、王治国、王学晶、邱玲、刘向祎、张秀明、黄宪章、李智、汪静。

引 言

检验程序的分析性能直接影响疾病诊断与监测的有效性。我国《医疗机构临床实验室管理办法》（卫医发〔2006〕73号）规定临床实验室应保证检验程序的完整性和有效性，国家标准GB/T 22576.1—2018《医学实验室——质量和能力的要求 第一部分：通用要求》要求临床实验室应选择经确认的检验程序，用前应验证所选程序的分析性能。

定量检验程序的分析性能指标有多种，包括测量准确度、测量正确度、测量精密度、测量不确定度、分析特异性（含分析干扰）、分析灵敏度、检测限和定量限及分析测量范围等。

定量检验程序分析性能的综合指标是准确度，可用测量不确定度描述，目前也常用总误差描述。关于总误差的意义和评估方法，可参阅WS/T 409。

在分析化学意义上，准确度由多种因素决定，主要包括精密度、正确度、特异性、分析测量范围内的线性等。目前临床检验中常见分析性能问题依检验项目或检验程序而异。精密度和正确度是基本分析性能问题，可见于各类检验项目。对某些原理（如化学、免疫学原理）的检验程序，特异性有时会是较突出的分析性能问题。对浓度跨度大的检验项目或某些原理的检验程序，有时可见线性不足问题。

基于以上分析，本标准给出精密度、正确度、特异性和线性验证的原则性方案设计和处理技术方法。

本标准尽量考虑了定量检验的多数情况，但部分方案对某些检验项目可能难以实现或不适用，在这种情况下，需参考有关文献酌情制定适宜验证方案。

临床实验室对检验程序的性能验证可能不限于上述四种性能指标，还有其他影响准确性的因素，如分析灵敏度、携带污染等，可参考有关文献酌情制定验证或评估方案。

定量检验程序分析性能验证指南

1 范围

本标准规定了定量检验程序分析性能（精密度、正确度、线性和特异性）验证的原则和方法。
本标准适用于医疗机构临床实验室对定量检验程序分析性能的验证。
本标准不适用于检验程序分析性能的建立或确认。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本标准必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本标准；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本标准。

GB/T 22576.1—2018 医学实验室——质量和能力的要求 第一部分：通用要求

GB/T 21415—2008 体外诊断医疗器械——生物样品中量的测量——校准品和控制物质赋值的计量学溯源性

WS/T 356 参考物质互换性评估指南

WS/T 403 临床化学检验常用项目分析质量标准

WS/T 406 临床血液检验常用项目分析质量标准

WS/T 409 临床定量检测方法分析总误差的评估

3 术语和定义

GB/T 21415—2008界定的以及下列术语和定义适用于本标准。

3.1

精密度 precision

在规定条件下，对同一或同类样品重复测量所得结果间的一致程度。

注1：精密度常用标准差或变异系数表示，是测量误差中的随机效应。

注2：规定条件在临床检验领域可以是批内、批间、实验室内等条件。

3.2

正确度 trueness

大批测量结果的均值与参考值间的一致程度。

注1：正确度常用偏倚表示，偏倚是均值与参考值的差值或其相对值，是测量误差中的系统效应。

注2：参考值可以是标准物质的定值、某种可靠方法的检测值或特定情况下的统计值。

3.3

互换性 commutability

由测量结果关系一致程度证明的参考物质的性质，测量结果关系一致程度是指，用两种指定测量程序测量参考物质所得结果的程序间关系，与两程序测量常规临床样品结果的程序间关系的一致程度。

注：互换性与参考物质的基质和分析物的性质有关，也与指定测量程序的原理和性能有关；不互换性反映参考物质基质和/或分析物与临床样品明显不同，也反映至少一种测量程序对分析干扰的易感性。

3.4

分析测量范围 analytical measurement range

在未对样品进行稀释、浓缩或其他非常规预处理的情况下，检验程序可直接检测的样品分析物浓度范围。

3.5

线性 linearity

在给定的分析测量范围内，检验程序使其检测结果与样品分析物浓度直接成比例的能力。

注1：检测结果是指最终结果，而非仪器输出的原始信号。

注2：非线性对分析测量范围内的单一浓度点表现为系统效应，对全部浓度点的影响可按随机效应处理。

3.6

干扰 interference

因样品其他成分或特性的影响而产生的检验结果变化。

注：干扰可见于检验程序非特异、指示反应欠佳、检验项目活性抑制等情况。

3.7

特异性 specificity

检验程序只检测目的检验项目的能力。

3.8

样品特定效应 sample-specific effect

干扰相关效应 interference-related effect

由未知干扰或非特异性造成的依样品而异的检验结果变化。

注：样品特定效应对单一样品为系统效应，对一组样品表现为随机效应。

4 总则

4.1 一般要求

临床实验室在启用新检验程序前应对其分析性能进行验证。实验室可能还有其他需验证分析性能的情况，如仪器搬迁或重要维修后，纠正重大环境设施失控情况后等。需验证的分析性能指标可能会因检验项目和实际情况而异，应酌情确定需验证的性能指标。

性能验证是指对已确认的检验程序的评价，对未经确认或自建检验程序，以及明显修改过的检验程序，应按相关要求进行分析性能确认。

临床实验室应根据临床需要的分析性能水平和目前技术水平制定检验程序分析性能标准[可接受的性能指标限值，一般包括随机误差（标准差或变异系数）和系统误差（偏倚）限值]，作为验证结果解释的主要依据。必要时，应制定不同条件下的性能标准。性能标准制定可参考有关标准、指南或文献的建议及检验程序制造或研发者的性能声明等。WS/T 403和WS/T 406分别给出常用生物化学和血液学检验项目分析性能（随机误差、系统误差和总误差）规范建议。

4.2 方案、程序和样品准备

本标准提出检验程序精密性、正确度、线性和特异性验证的原则性方案，不排除其他合理方案。各项性能验证可独立进行，也可根据各项验证方案要点，酌情合并进行，以节省时间和资源。

在进行验证实验前，实验室工作人员应充分理解检验程序，熟悉相关操作，制定内部质量控制方法和规则，保证检验程序正常运行。每批验证实验均应进行内部质量控制，若出现失控情况，应弃去该批实验结果，分析可能原因，重新进行实验。

实验前应根据方案设计对样品进行必要处理，如复溶（对于冻干样品）、混和及分装等，分装份数应满足各批（包括可能额外检测批）检测需要。对于部分检验程序，样品可能还需其他特定处理，如沉淀、提取等，在此情况下，应独立进行各批检测的样品处理。若需储存或运输样品，应根据检验项目性质，确定合适的样品储存和运输条件，保证样品在验证实验期间稳定完整。

4.3 数据处理

在进行数据处理前，应检查实验结果中的可能离群值，可用特定统计规则鉴别并剔除离群值，但剔除数据量不得超过总数据量的5%。

本标准数据处理中随机效应和系统效应分别用标准差和偏倚或差值描述，进行相关计算时也可用其相对值，即变异系数和相对偏倚或差值，将标准差和偏倚或差值除以对应的均值得变异系数和相对偏倚或差值。

5 精密度验证

5.1 方案设计

精密度验证的基本过程是用待验证程序对多水平样品进行多批、多次检测，获得检验程序的重复性（批内）和中间（实验室内）精密度指标（标准差或变异系数），与规定的可接受标准比较，得出验证结论。

精密度验证实验应检测至少2个不同浓度水平的样品，对样品进行至少5批检测，每批在不同工作日（不一定连续）完成，每批重复检测每种样品至少3次。

5.2 样品准备与检测

按上述方案准备样品，宜选用单一或混合患者样品，也可选用性质相似的质控样品；样品水平应接近医学决定水平或相关病理情况下的临床常见水平，避免极端水平。

按上述方案用待验证检验程序检测样品，记录检测结果。

5.3 数据处理

可用一般办公软件的相关函数进行数据处理。计算每个样品每批实验结果的均值和标准差(s_{WRi})，计算各批均值的均值（总均值）和标准差(s_M)；按下列公式计算平均批内标准差(s_{WR})、批间标准差(s_{BR})和实验室内标准差(s_{WL})：

$$s_{WR} = \sqrt{\frac{\sum_{i=1}^{n_1} s_{WRi}^2}{n_1}} \dots\dots\dots (1)$$

$$s_{BR} = \sqrt{s_M^2 - \frac{s_{WR}^2}{n_2}} \dots\dots\dots (2)$$

$$s_{WL} = \sqrt{s_{WR}^2 + s_{BR}^2} = \sqrt{\frac{n_2-1}{n_2} s_{WR}^2 + s_M^2} \dots\dots\dots (3)$$

式中：

n_1 ——实验批数；

n_2 ——每批实验的重复检测数。

在 s_{WL} 大于实验室规定标准差(s_0)情况下，可考虑检验差异的统计学显著性，可用卡方(χ^2)检验，按下式计算 χ^2 值：

$$\chi^2 = \nu \left(\frac{s_{WL}}{s_0} \right)^2 \dots\dots\dots (4)$$

式中：

ν ——自由度。

s_{WL} 由 s_{WR} 和 s_M 合成，其有效自由度可由Welch-Satterthwaite公式估计：

$$\nu = \frac{s_{WL}^4}{\frac{(\frac{n_2-1}{n_2})^2 s_{WR}^4}{n_1(n_2-1)} + \frac{s_M^4}{n_1-1}} \dots\dots\dots (5)$$

查 χ^2 表或用办公软件相关函数计算特定假排除概率（如0.05）下与上述自由度对应的 χ^2 界值，若实验 χ^2 值大于理论 χ^2 界值，则可认为实验标准差明显大于规定标准差，反之亦然。

5.4 结果解释

一般情况下，若 s_{WL} 小于或接近实验室规定的分析不精密度标准，认为该检验程序精密度可接受。若 s_{WL} 大于规定标准，可考虑进行上述统计学检验，若 s_{WL} 明显大于规定标准，提示分析不精密度不可接受。必要时，还可结合其他随机误差（非线性和非特异性）验证结果，作综合分析。

6 正确度验证

6.1 方案设计

正确度验证可采用参考物质检测和程序对比等方式，通过实验获得待验证检验程序的偏倚，与规定的可接受偏倚比较，得出验证结论。

若采用参考物质检测方式，正确度验证应检测至少2个浓度水平的参考物质，对每种参考物质进行至少10次重复检测；若采用程序对比方式，应检测至少20份患者样品，每份样品用每种程序检测1次。两种方式的检测均可在一批或多批实验中完成。

检验程序的正确度直接影响医学决定水平或参考区间的有效性，正确度验证方式选择及参考物质或对比方法选择应充分考虑相关检验项目的标准化情况，合理制定验证方案。

目前，部分检验项目标准化程度较高，不同检验程序在新鲜患者样品上结果接近，在临床上使用相同医学决定水平或参考区间，这些检验项目一般具有公认的参考测量程序，有的还具有定值可溯源且可互换的参考物质，本标准称此类检验项目为A类。对于此类检验项目，正确度验证的目的是保证检验结果的准确性，从而使通用医学决定水平或参考区间得以有效实施。

还有许多检验项目标准化程度不高，不同检验程序结果存在差异，使用不同医学决定水平或参考区间，这些检验项目一般无公认的参考测量程序，也无通用的参考物质，本标准称此类检验项目为B类。对于此类检验项目，正确度验证的目的是保证检验结果在特定检验程序范围内的可比性，从而使特定医学决定水平或参考区间得以有效实施。

采用参考物质检测方式时还应充分考虑参考物质的互换性。互换性的意义及评价方法见WS/T 356。

6.2 通过参考物质检测验证正确度

6.2.1 参考物质选择、准备与检测

可酌情依次选用下列种类的参考物质：

- 1) 可互换的有证参考物质：一般来自参考物质研制机构，适用A类检验项目；
- 2) 可互换的正确度控制物质：应符合GB/T 21415相关要求；一般来自室间质评机构或标准化计划组织者，也可由实验室自行制备，适用A类检验项目；
- 3) 专用正确度控制物质：应符合GB/T 21415相关要求，且适用待验证检验程序；此类物质可能不具互换性，但对特定检验程序组有可靠定值，有的称校准验证物质，也可以是其他符合要求的物质，一般由检验程序制造商或其他机构提供，也可由实验室自行制备，适用A类和B类检验项目。

所选参考物质浓度水平应接近医学决定水平或参考区间重要限值，或相关病理情况下的临床常见水平，避免极端水平。

按上述方案准备足够数量的参考物质。

按上述方案用待验证检验程序检测参考物质，记录检测结果。

6.2.2 数据处理

记录或计算参考物质定值 (c) 和定值的标准不确定度 (u)，上述第1、2、3类参考物质一般会给出定值和标准不确定度或扩展不确定度及包含因子，将扩展不确定度除以包含因子可得标准不确定度；第4类参考物质定值为同程序组均值，其标准不确定度为该均值的标准误，即实验室间标准差除以参加实验室数的平方根。

计算各检测值的均值 (m) 和标准差 (s)，按下列公式计算偏倚 (b) 及其标准差 (s_b)：

$$b = m - c \dots\dots\dots (6)$$

$$s_b = \sqrt{\frac{s^2}{n} + u^2} \dots\dots\dots (7)$$

式中：

n ——重复检测数。

偏倚的统计学显著性检验可用简化的 t 检验，若偏倚的绝对值 $|b| > 2s_b$ ，则认为存在显著偏倚。

6.2.3 结果解释

一般情况下，若 $|b|$ 小于或接近实验室规定的偏倚，认为该检验程序正确度可接受；若出现 $|b| > 2s_b$ 的情况，提示检验程序存在显著偏倚，但认为此偏倚在临床上可接受。

若 $|b|$ 大于实验室规定的偏倚，且大于 $2s_b$ ，提示检验程序正确度不可接受；若 $|b|$ 大于实验室规定偏倚，但小于 $2s_b$ ，提示检验程序精密度不足或所选参考物质不确定度太大，应分析原因，酌情考虑重新实验。

6.3 通过程序对比验证正确度

6.3.1 对比程序选择与准备

可酌情选用下列对比程序：

- 1) 参考测量程序：适用A类检验项目；
- 2) 医学决定水平或参考区间相同的其他常规检验程序：适用A类和B类检验项目，应选择应用相对广泛、已知分析性能良好的程序。

对比程序可由本实验室运行，也可由其他实验室运行，实验前相关实验室均应熟悉验证方案，准备对比程序，保证程序正常运行。

6.3.2 样品准备与检测

按上述方案准备足够数目和体积的患者样品，浓度水平包括医学决定水平或参考区间重要限值及相关病理情况下的临床常见水平，避免极端水平。将每份样品一分为二，分别供待验证程序 and 对比程序检测用。

样品宜新鲜检测，必要时可做储存。

按上述方案用待验证检验程序和对比程序检测样品，记录检测结果。

6.3.3 数据处理

计算每个样品的待验证程序检测值与对比程序检测值之差，将差值对样品浓度作图，检查差值有无趋势性变化，若有，应酌情分段处理数据。

计算差值的均值和标准差，即为检验程序的偏倚 (b) 及其标准差 (s_b)。偏倚的统计学显著性检验见本标准第6.2.2条。

6.3.4 结果解释

参见本标准第6.2.3条。在 $|b|$ 大于实验室规定偏倚但小于 $2s_b$ 情况下，提示两种检验程序中的一种或两种精密度不足或存在明显的样品特定效应（非特异性），应分析原因，酌情考虑重新实验。

7 线性验证

7.1 方案设计

线性验证的基本过程是用待验证检验程序检测覆盖特定浓度范围的已知浓度的多个样品,将检测浓度对已知浓度进行线性回归,计算特定回归参数,与规定的相关标准比较,得出验证结论。

线性验证应检测至少5个浓度的样品,最高和最低浓度应分别接近检验程序分析测量范围的上下限,每个样品重复检测至少3次,检测可在一批实验中完成。

7.2 样品准备与检测

线性验证宜选用患者样品,也可选用其他适宜样品。理论上,样品可以是不同浓度的自然样品(如来自不同个体的样品),也可以是由2个高低浓度样品按不同比例混合而制备的系列混合样品。前者需用另种检验程序为样品定值,影响实验结果的因素更多,结果解释更为复杂,故线性验证一般考虑选用系列混合样品,但对于不便制备混合样品的检验项目(如血细胞分析),可能适宜选用患者样品。

若选用混合样品,准备高低浓度两份样品,其浓度分别接近待验证程序分析测量范围的上下限,将两份样品按不同比例混合,得不同浓度样品,由高低值样品浓度和混合比例计算样品浓度,做已知浓度。高低值样品浓度是本程序检测值。应合理确定混合比例,使样品浓度等距或接近等距。样品混合一般采用容量法,应合理选用移液器具和移液体积,保证所转移样品体积的准确性(相对于程序的精密度,混合体积误差可以忽略)。需要时,可考虑重量法混合。高低值样品难以获得时,可考虑添加分析物或稀释样品,应合理选用添加物和稀释剂,尽量减小可能的样品互换性改变。

若选用患者样品,其浓度应覆盖该程序分析测量范围,尽量接近等距分布,用另种检验程序为样品定值,做已知浓度。

按上述方案用待验证程序检测样品,记录检测结果。

7.3 数据处理

以已知浓度做自变量(x),各单次检测浓度做因变量(y),进行直线回归,得回归方程:

$$y = b_1x + b_0 \dots\dots\dots (8)$$

用回归方程计算各已知浓度下的预测值($\hat{y}$),用下式计算直线回归的剩余标准差(s):

$$s_{y|x} = \sqrt{\frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (y_{ij} - \hat{y}_i)^2}{n_1 n_2 - 2}} \dots\dots\dots (9)$$

式中:

n_1 ——样品数;

n_2 ——每样品的重复检测数。

按本标准第5.3条所述方法计算批内标准差(s_{WR}),在 $s_{y|x} > s_{WR}$ 情况下,可用 F 检验判断 $s_{y|x}$ 是否显著大于 s_{WR} (有无显著非线性),用下列公式计算 F 值及 $s_{y|x}$ 和 s_{WR} 的自由度($\nu_{y|x}$ 和 ν_{WR}):

$$F = \frac{s_{y|x}^2}{s_{WR}^2} \dots\dots\dots (10)$$

$$\nu_{y|x} = n_1 n_2 - 2 \dots\dots\dots (11)$$

$$\nu_{WR} = n_1(n_2 - 1) \dots\dots\dots (12)$$

查 F 表或用办公软件相关函数计算特定假排除概率(如0.05)下与上述自由度对应的 F 值,若 F 值大于临界 F 值,则认为 $s_{y|x}$ 显著大于 s_{WR} 。

在 $s_{y|x}$ 显著大于 s_{WR} 情况下，可按式计算非线性标准差（ s_{NL} ）：

$$s_{NL} = \sqrt{s_{y|x}^2 - s_{WR}^2} \dots\dots\dots (13)$$

7.4 结果解释

一般情况下，若 $s_{y|x}$ 不显著大于 s_{WR} ，认为检验程序的线性可接受；若 $s_{y|x}$ 显著大于 s_{WR} ，但 s_{NL} 小于实验室规定的非线性标准差，认为此非线性在临床上可接受。

若 $s_{y|x}$ 显著大于 s_{WR} ，且 s_{NL} 显著大于实验室规定的非线性标准差，提示检验程序线性不可接受。必要时，还可结合其他随机误差（分析不精密度和非特异性）验证结果，作综合分析。

8 特异性验证

8.1 方案设计

特异性验证可采用干扰试验或程序对比等方式。干扰试验的基本过程是用待验证程序检测基础样品和添加可能干扰物的基础样品，比较两种样品的检测结果，判断有无明显干扰；程序对比的基本过程是用待验证程序和另一种已知性能良好的检验程序同时检测多个代表性患者样品，通过方差分析估计样品特定效应（非特异性），与规定的相关指标比较，得出验证结论。

干扰试验可明确判断已知可能干扰物对检验结果有无明显影响，但不能确定未试验干扰物及未知干扰因素的可能影响；程序对比可估计实验样品代表的所有可能因素的影响，但不能确定影响因素的性质。对部分测量原理的检验程序，可能影响因素众多，无从得知具体干扰物，程序对比方式给出更具实际意义的特异性信息。

干扰试验应重复检测基础样品和添加样品至少10次；程序对比应检测至少20患者样品，每样品重复检测至少2次。两种方式的验证均可在一批实验中完成。

8.2 通过干扰试验验证特异性

8.2.1 样品准备与检测

干扰试验基础样品宜选用单一或混合正常患者样品，也可选用其他适宜样品，浓度应接近医学决定水平或重要参考区间限值。应准备足够体积的样品，满足基础样品和添加样品配制的需要。

可能干扰物一般包括内源性物质（如异常水平的胆红素、血红蛋白、甘油三酯、特定蛋白质、特定代谢物以及正常或异常水平的与分析物结构相似的物质）、外源性物质（如药物、食品成分）及添加物（如抗凝剂）等，可根据检验程序的测量原理及其已知可能干扰因素，酌情选择试验干扰物。

应独立制备每种选定干扰物的添加样品。向基础样品中添加干扰物，制备添加样品。干扰物浓度一般选用临床样品常见的较高浓度（如医院患者人群的90%分位浓度附近）。可先配制或选用干扰物原液，再将原液加入基础样品。应合理确定原液溶剂和浓度或选用合适干扰物制剂，保证干扰物在样品中完全溶解，样品基质不发生明显变化。添加原液时，应向基础样品中加入相同比例的原液溶剂。

按上述方案用待验证程序检测基础样品和添加样品，记录检测结果。

8.2.2 数据处理

计算基础样品和添加样品的均值（ c_0 和 c_+ ）和标准差（ s_0 和 s_+ ），按下列公式计算两样品差异（ d ）及其标准差（ s_d ）：

$$d = c_+ - c_0 \dots\dots\dots (14)$$

$$s_d = \sqrt{\frac{s_0^2 + s_+^2}{n}} \dots\dots\dots (15)$$

式中：

n ——重复检测数。

两样品差异的统计学显著性检验可用简化的 t 检验, 若 $|d| > 2s_d$, 则认为存在显著差异。

8.2.3 结果解释

结合正确度验证结果和其他干扰实验结果判断验证结果。计算总偏倚, 若总偏倚小于或接近实验室规定的允许偏倚, 则认为干扰可接受, 若 $|d| > 2s_d$, 提示存在统计学显著干扰, 但认为此干扰在临床上可接受。

若总偏倚大于实验室规定的偏倚, 且 $|d| > 2s_d$, 提示干扰不可接受。

8.3 通过程序对比验证特异性

8.3.1 对比程序选择与准备

特异性验证理想的对比程序为参考测量程序, 不适用或不可及时, 可选择应用相对广泛、已知分析性能良好的其他常规检验程序。

对比程序可由本实验室运行, 也可由其他实验室运行, 实验前相关实验室均应熟悉验证方案, 准备对比程序, 保证程序正常运行。

8.3.2 样品准备与检测

按上述方案准备足够数目和体积的来自不同个体的患者样品, 样品应有代表性, 浓度水平包括医学决定水平或参考区间重要限值及相关病理情况下的临床常见水平, 避免极端水平。将每份样品一分为二, 分别供待验证程序 and 对比程序检测用。

样品宜新鲜检测, 必要时可做储存。

按上述方案用待验证检验程序和对比程序检测样品, 记录检测结果。

8.3.3 数据处理

按本标准第5.3条所述方法计算待验证程序和对比程序的批内标准差 (s_{WR1} 和 s_{WR2}), 计算每个样品的每个程序结果的均值, 计算每个样品两程序均值的差值 (d), 计算全部样品 d 的标准差 (s_d), 按下式计算 d 的精密度相关标准差 (s_{PR}):

$$s_{PR} = \sqrt{\frac{s_{WR1}^2 + s_{WR2}^2}{n}} \dots\dots\dots (16)$$

式中:

n ——每个样品每种程序的重复检测数。

若 $s_d > s_{PR}$, 可按本标准第7.3条所述方法检验差异是否显著, 并按下列公式计算样品特定效应标准差 (s_{SS}):

$$s_{SS} = \sqrt{s_d^2 - s_{PR}^2} \dots\dots\dots (17)$$

8.3.4 结果解释

一般情况下, 若 s_d 接近 s_{PR} , 或 s_{SS} 小于或接近实验室规定的样品特定效应标准差, 认为检验程序的特异性可接受。

若 s_{SS} 大于实验室规定的样品特定效应标准差, 且 s_d 明显大于 s_{PR} , 当对比程序为参考测量程序时, 提示检验程序存在不可接受的样品特定效应; 当对比程序为常规检验程序时, 提示待验证程序或 (和) 对比程序存在明显分析干扰, 可分析可能原因, 酌情选用另种对比程序重新实验。必要时, 还可结合其他随机误差 (分析不精密度和非线性) 验证结果, 作综合分析。

附录 A (资料性) 分析性能验证示例

A.1 引言

以高密度脂蛋白胆固醇(HDL-C)检验为例,说明检验程序分析性能验证过程。本示例基于一项HDL-C方法评价研究(中华检验医学杂志 2009; 32: 521),为便于说明计算方法,采用模拟数据。本示例得出的分析性能情况(如不精密度、非线性、非特异性等表现)仅反映HDL-C的可能情况,其他类别检验项目可能会有不同性能表现。本示例所列各项验证,仅以示例为目的,实际工作中宜根据检验项目性质和应用及实验室条件等,酌情选择适宜的验证实验及实验方式。

本示例假设如下场景:实验室一直使用程序A进行HDL-C检验,该程序运行情况良好,分析性能符合相关要求,但由于某种原因,实验室拟改用程序B进行HDL-C检验,现对程序B进行分析性能验证。

根据有关标准和指南,实验室规定HDL-C允许总随机误差4%,允许系统误差5%;根据工作经验、文献报告和试剂说明书等,实验室规定允许实验室内分析不精密度、非线性标准差和样品特定效应标准差各为2%(三者的合并效应约3.5%),以保证总随机变异小于4%。

A.2 精密度验证

选用混合临床标本血清做实验样品,收集HDL-C接近医学决定水平(1.03 mmol/L)和临床常见水平(>1.5 mmol/L)的剩余血清,混匀,分装5份,密封, -70 °C 保存。用程序B对样品进行5批检测,每批重复检测3次。

用Excel办公软件处理数据(见附件电子表格),样品1处理过程如下:

将5批实验结果列于电子表格B2~D6,用AVERAGE和STDEV函数计算批均值和批内标准差 s_{WRi} (E2~E6、F2~F6),计算 s_{WRi}^2 (G2~G6);

本实验 n_1 和 n_2 分别为5和3,列于E7、E8,用上述函数计算批均值的均值(总均值)和标准差(s_M)(E9、E10),按公式(1)和(3)计算平均批内标准差(s_{WR})和实验室内标准差(s_{WL})(E11、E12), s_{WL} 计算结果为0.0242 mmol/L;

实验室规定HDL-C分析不精密度为2%,折算为标准差(s_0)为0.0208 mmol/L(E13), s_{WL} 略大于 s_0 ,进行 χ^2 检验;

按公式(5)计算自由度 ν (E14),按公式(4)计算 χ^2 值(E15,结果为17.6),用CHIINV函数计算0.05假排除概率下的临界 χ^2 值(χ^2_{crit})(E16,结果为22.4), $\chi^2 < \chi^2_{crit}$,故 s_{WL} 不明显高于 s_0 。

对水平2样品数据做相同处理,得总均值1.55 mmol/L, s_{WL} 0.0263 mmol/L, s_0 0.0310 mmol/L; $s_{WL} < s_0$,无须进行 χ^2 检验。

根据上述实验和分析,认为程序B分析不精密度可接受。

A.3 正确度验证

A.3.1 通过参考物质检测验证正确度

选用国家一级标准物质GBW09193和GBW09195(冰冻血清)做正确度验证物质,两种标准物质的定值(c)分别为1.49 mmol/L和1.02 mmol/L,定值的扩展不确定度(U)均为0.04 mmol/L(包含因子为2)。用程序B在同批实验中检测标准物质10次。

用Excel办公软件处理数据(见附件电子表格),GBW09193处理过程如下:

将10次检测结果列于电子表格B2~K2,用AVERAGE和STDEV函数计算均值(m)和标准差(s)(L2、M2);

本实验 n 为10,列于M3,标准物质定值(c)及其扩展不确定度(U)分别列于M4、M5,计算标准物质定值的标准不确定度(u)(M6),按公式(6)和(7)计算偏倚(b)及其标准差(s_b)(M8、M9), b 的计算结果为0.038 mmol/L;

实验室规定允许偏倚为5%,折算为绝对偏倚(b_0)为0.075 mmol/L(L7)。

对GBW09195数据做相同处理,得 b 0.030 mmol/L, s_b 0.0205 mmol/L, b_0 0.051 mmol/L。

两种标准物质 b 均 $< b_0$,且 $< 2s_b$,认为程序B偏倚可接受。

A.3.2 通过程序对比验证正确度

选用程序A做对比程序。收集HDL-C 1 mmol/L~1.7 mmol/L临床标本血清20份，用程序A和程序B检测HDL-C，每份样品用每种程序检测1次。

用Excel办公软件处理数据（见附件电子表格），处理过程如下：

将两种程序测定20份样品结果列于电子表格B2~C21，计算每份样品两程序结果的差值（程序B-程序A），将差值对程序A结果做图（见附件），差值及其幅度在本实验浓度范围内未见明显变化趋势，故合并处理20数据；

用AVERAGE和STDEV函数计算差值的均值（ b ）和标准差（ s_b ）（D24、D25）， b 和 s_b 计算结果分别为0.032 mmol/L和0.046 mmol/L；

计算程序A检测结果均值（ m ）（B22），实验室规定允许偏倚为5%，折算为绝对偏倚（ b_0 ）为0.067 mmol/L（M7）。

$b < b_0$ ，且 $< 2s_b$ ，认为程序B的偏倚可接受。

A.4 线性验证

选用5水平混合样品（由低到高，样品1~5）进行线性验证，收集HDL-C < 1 mmol/L和 > 2 mmol/L临床标本血清，混匀，得样品1和5，精密等体积混合样品1和5，得样品3，同法混合样品1和3及3和5，分别得样品2和4。用程序B检测5种样品，每种样品重复检测3次。

用Excel办公软件处理数据（见附件电子表格），处理过程如下：

将检测结果分3组列于电子表格C2~C16，用AVERAGE函数计算水平1和5的3次检测的均值，做其已知值（D2、D6），按稀释比例计算水平2~3的已知值（D3~D5），将5水平已知值（D2~D6）粘贴于另两组数据的相应位置（D7~D11及D12~D16）；

选定B17~C21区域，用LINEST函数将检测值对已知值进行直线回归，得剩余标准差（ $s_{y|x}$ ）（C19）、自由度（ $\nu_{y|x}$ ）（C20）等回归参数， $s_{y|x}$ 结果为0.0318；

按A2.1所述方法计算平均批内标准差（ s_{WR} ）（G2）及其自由度（ ν_{WR} ）（H2）， s_{WR} 计算结果为0.0186， $s_{y|x}$ 似明显大于 s_{WR} ；

用F检验比较 $s_{y|x}$ 和 s_{WR} ，按公式（10）计算F值（B22），用FINV函数计算0.05假排除概率下的F临界值（ F_{crit} ），计算结果分别为2.913和2.887， $F > F_{crit}$ ，提示程序B可能存在明显非线性；

将检测值（均值）对已知值作图（见附件），图中实线为 $y=x$ 直线，虚线为检测值拟合线，可见检测值表现一定程度的似呈二阶曲线形式的非线性；

按公式（13）计算非线性标准差（ s_{NL} ）（D25），结果为0.0258 mmol/L；

计算样品已知值的均值（ m ）（D24），实验室规定允许非线性标准差为2%，折算为绝对标准差（ s_{NL0} ）为0.0321 mmol/L（D26）。

$s_{NL} < s_{NL0}$ ，认为程序B的非线性可接受。

A.5 特异性验证

A.5.1 通过干扰实验验证特异性

根据检验项目性质、方法原理等，认为影响HDL-C结果的主要因素可能是血清中的其他脂蛋白，尤其是极低密度脂蛋白（VLDL），拟试验VLDL对程序B的干扰。VLDL是血清甘油三脂（TG）的主要来源，临床标本中常见的较高TG水平为3 mmol/L~5 mmol/L，收集此浓度范围的临床标本血清，超速离心法分离VLDL，吸取离心管顶部约10%体积的VLDL组分，做干扰物原液；另收集TG < 1.7 mmol/L的临床标本血清，做基础样品；取基础样品2份，分别精密加入10%体积的干扰物原液或生理盐水，混匀；用程序B测定基础样品和添加样品HDL-C各10次。

用Excel办公软件处理数据（见附件电子表格），处理过程如下：

将基础样品和添加样品检测结果分列于电子表格B2~B11和C2~C11，用AVERAGE函数计算其均值（B13、C13），用STDEV函数计算其标准差（B14、C14），用公式（13）和（14）计算两样品均值的差值（ d ）及其标准差（ s_d ）（C15、C16），结果分别为0.0150 mmol/L和0.0057 mmol/L， $d > 2s_d$ ，提示较高水平VLDL可能对程序B有明显影响，可使HDL-C结果升高约1.2%；

正确度验证结果显示程序B偏倚小于3% (A3)，实验室规定允许偏倚为5%，其他干扰试验未见明显影响（假设），认为VLDL对程序B的影响可接受。

A. 5.2 通过程序对比验证特异性

选用程序A做对比程序。收集HDL-C 1 mmol/L~1.7 mmol/L临床标本血清20份，用程序A和程序B检测HDL-C，每份样品用每种程序检测2次。

用Excel办公软件处理数据（见附件电子表格），处理过程如下：

将两种程序测定20份样品结果列于电子表格B3~C22和G3~H22，用AVERAGE函数计算各程序各样品的均值 ($\bar{m}$) (D3~D22、I3~I22)，用STDEV函数计算批内标准差 (s_{WRi}) (E3~E22、J3~J22)，计算 s_{WRi} 的平方 (F3~F22、K3~K22)，计算两程序均值的差值 (d) (L3~L22)，计算平均批内标准差 (s_{WR}) (E23、J23)；

用STDEV函数计算差值的标准差 (s_d) (L24)，结果为0.0307 mmol/L，用公式 (15) 计算精密度相关标准差 (s_{PR}) (L25)，结果为0.0139 mmol/L，经检验， s_d 明显大于 s_{PR} ，用公式 (16) 计算样品特定效应标准差 (s_{SS}) (L26)，结果为0.0274 mmol/L。

实验室规定允许样品特定效应标准差为2%，折算为绝对值为0.0275 mmol/L， s_{SS} 与之接近，认为程序B的样品特定效应标准差可接受。

参 考 文 献

- [1] 国家质量监督检验检疫总局. 标准物质的选择与应用: JJF 1507—2015. 2015.
 - [2] 国家质量监督检验检疫总局. 测量不确定度评定与表示: JJF 1059.1—2012. 2012.
 - [3] 国家质量监督检验检疫总局. 标准物质定值的通用原则及统计学原理: JJF 1343—2012. 2012.
 - [4] Killeen AA, Long T, Souers R, et al. Verifying Performance Characteristics of Quantitative Analytical Systems: Calibration Verification, Linearity, and Analytical Measurement Range. *Arch Pathol Lab Med*, 2014,138:1173–1181.
 - [5] Krouwer JS. Setting performance goals and evaluating total analytical error for diagnostic assays. *Clin Chem*, 2002,48:919–927.
 - [6] Miller WG, Schimmel H, Rej R, et al. IFCC Working Group Recommendations for Assessing Commutability Part 1: General Experimental Design. *Clin Chem*, 2018,64:447–454.
 - [7] Nilsson G, Budd JR, Greenberg N, et al. IFCC Working Group Recommendations for Assessing Commutability Part 2: Using the Difference in Bias between a Reference Material and Clinical Samples. *Clin Chem*, 2018,64:455–464.
 - [8] Stöckl D, Dewitte K, Thienpont LM. Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data? *Clin Chem*, 1998,44:2340–2346.
-